Hitchhikers' guide to **mixed effects regression models** **Logistic ME regression**

GLMER

International school in psycholinguistics, neurolinguistics and clinical linguistics

Dr. Seçkin Arslan

seckin1984@gmail.com

@seckin1984

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

26.10.21

This lecture

This lecture:

(i) **Generalized linear mixed-effects regression (glmer)**

(ii) **Example with accuracy data**

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

This lecture

When we fit a linear regression model, we always assume the response has a continuous scale (e.g. response times)

What if we analyse ordinary response on 1-5 scale ?

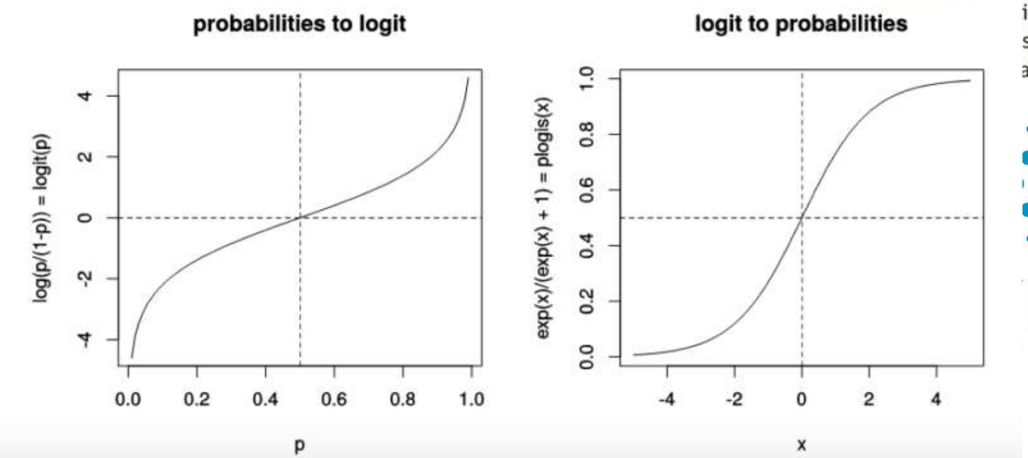
We cannot use linear lmm models for binary variables 1-0
This is more of a classification problem.

Hitchhikers' guide to **mixed effects regression models**

- **Generalized linear (mixed-effects) regression = GLM(ER)**

  is a generalization of linear regression

  - Logistic regression = dependent variable is binary (1 = accurate , 0 = inaccurate)
  - Logistic regression converts binary data into continuous data via "log odds link" function.



ional school in
s, neurolinguistics and
al linguistics

ERSITÉ
D'AZUR

Hitchhikers' guide to **mixed effects regression models**

- **We will use glmer (generalized mixed-effects regression) to analyse proportions of looks**
  - Logistic regression automatically transforms binary data with log odds link function **log(p / (1-p))**
  - this represents probability in a form of continuous value.
  - Probability in this case is whether participants looks at target (1) or non-target (0).

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models**

- **Logistic mixed-effects regression assumes**
  - Relationship between dependent variable and independent variable is linear.
  - Multicollinearity is not strong
  - And there are no assumptions about normality of distribution of residuals.

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models**

- **How to import your data into R**

COGNITIVE NEUROPSYCHOLOGY, 2017
https://doi.org/10.1080/02643294.2017.1394284

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Predicting the sources of impaired *wh*-question comprehension in non-fluent aphasia: A cross-linguistic machine learning study on Turkish and German

Seçkin Arslan[a], Eren Gür[b] and Claudia Felser[a]

[a]Potsdam Research Institute for Multilingualism, University of Potsdam, Potsdam, Germany; [b]Department of Neurology, Hamidiye Şişli Etfal Research and Training Hospital, Şişli, Istanbul, Turkey

ABSTRACT
This study investigates the comprehension of *wh*-questions in individuals with aphasia (IWA) speaking Turkish, a non-*wh*-movement language, and German, a *wh*-movement language. We examined six German-speaking and 11 Turkish-speaking IWA using picture-pointing tasks. Findings from our experiments show that the Turkish IWA responded more accurately to both object *who* and object *which* questions than to subject questions, while the German IWA performed better for subject *which* questions than in all other conditions. Using random forest models, a machine learning technique used in tree-structured classification, on the individual data revealed that both the Turkish and German IWA's response accuracy is largely predicted by the presence of overt and unambiguous case marking. We discuss our results with regard to different theoretical approaches to the comprehension of *wh*-questions in aphasia.

ARTICLE HISTORY
Received 26 January 2017
Revised 28 July 2017
Accepted 13 October 2017

KEYWORDS
Non-fluent aphasia; random forest algorithm; sentence comprehension; wh-in-situ; wh-questions; wh-movement

The data is from this paper, you can read if you are interested.

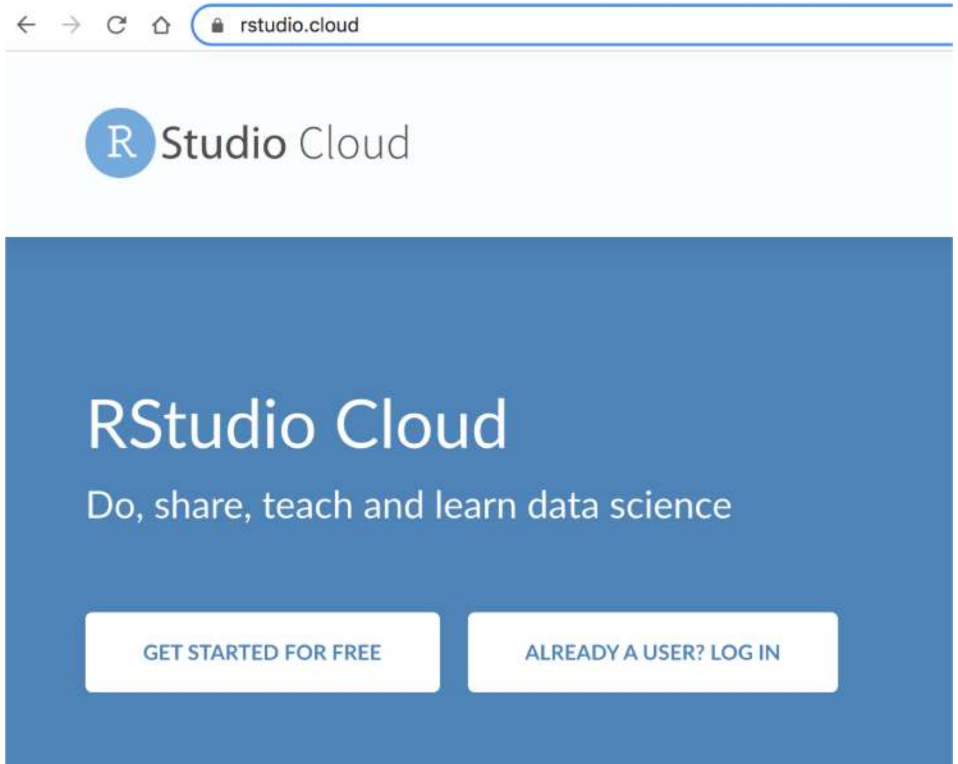
International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models**

- **Follow from Rstudio Cloud (Recommended)**

https://rstudio.cloud/



International school in psycholinguistics, neurolinguistics and clinical linguistics

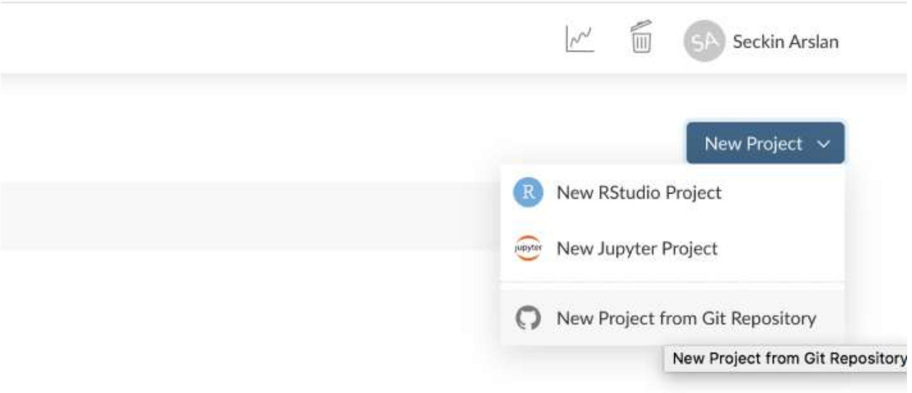
UNIVERSITÉ CÔTE D'AZUR

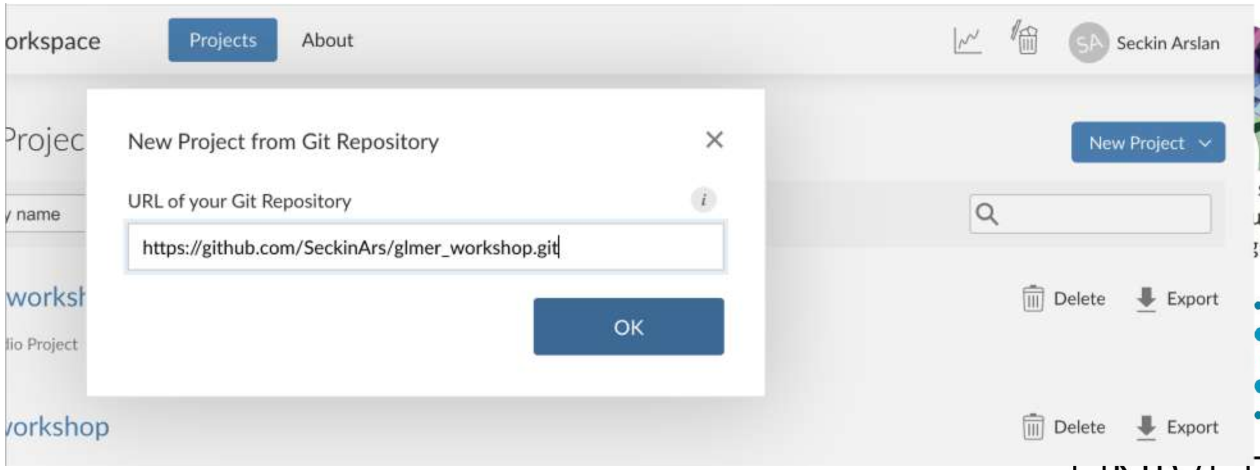
Hitchhikers’ guide to **mixed effects regression models**

- **Follow from Rstudio Cloud (Recommended)**

https://rstudio.cloud/

https://github.com/SeckinArs/glmer_workshop.git

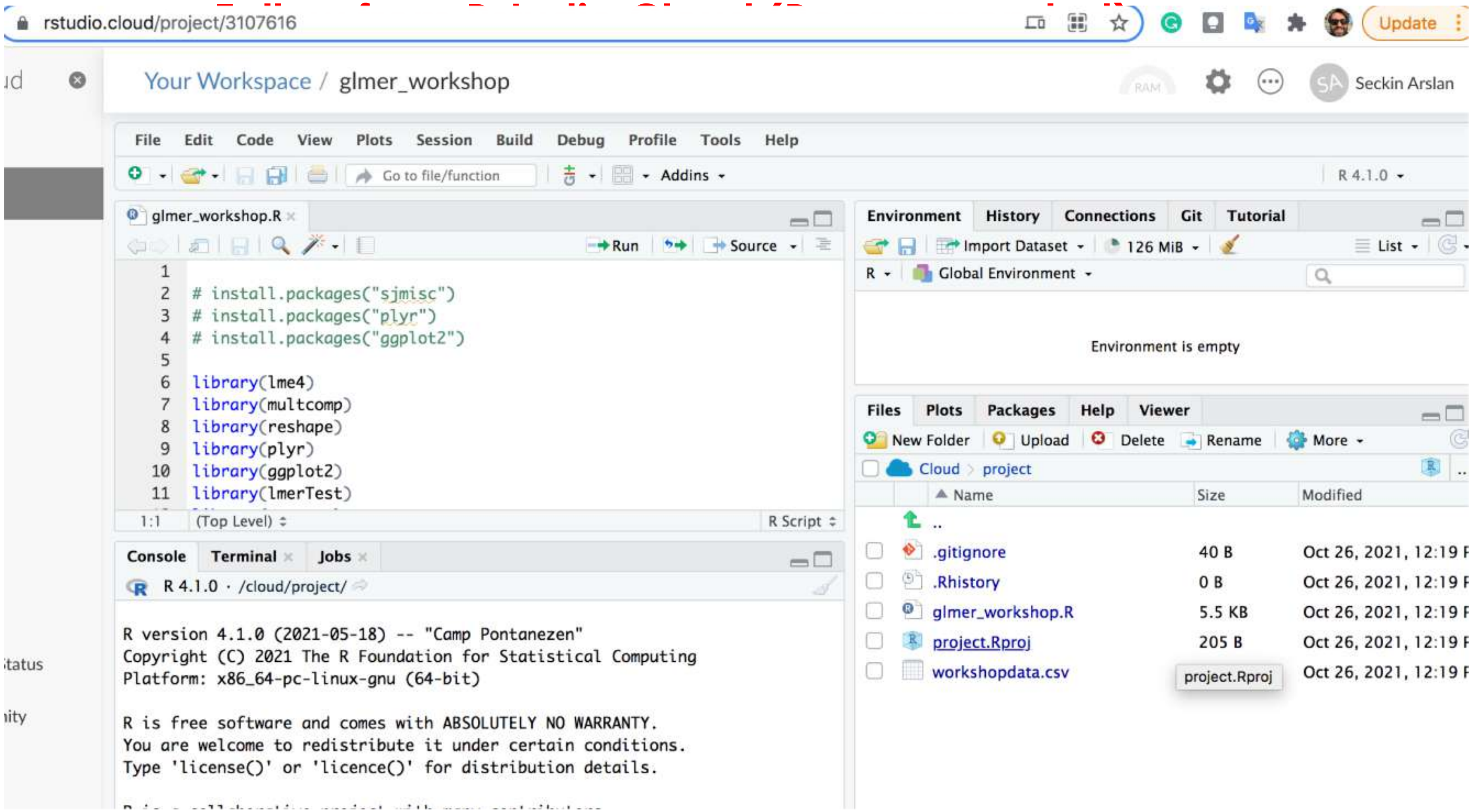




school in
urolinguistics and
guistics

UNIVERSITÉ
CÔTE D’AZUR

Hitchhikers' guide to **mixed effects regression models**

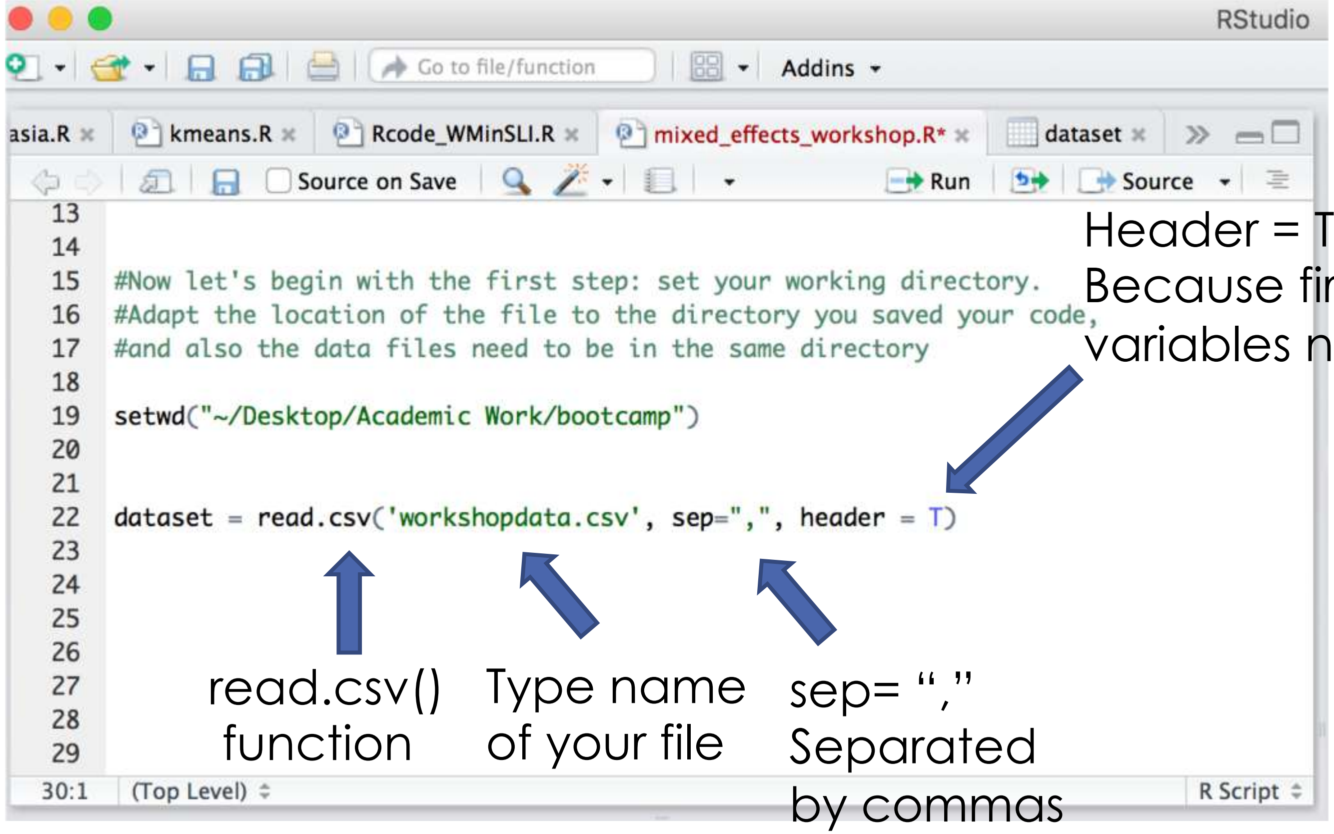


International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models**

How to import your data into R



Header = True
Because first row has variables names

read.csv() function

Type name of your file

sep= "," Separated by commas

If this does not work!

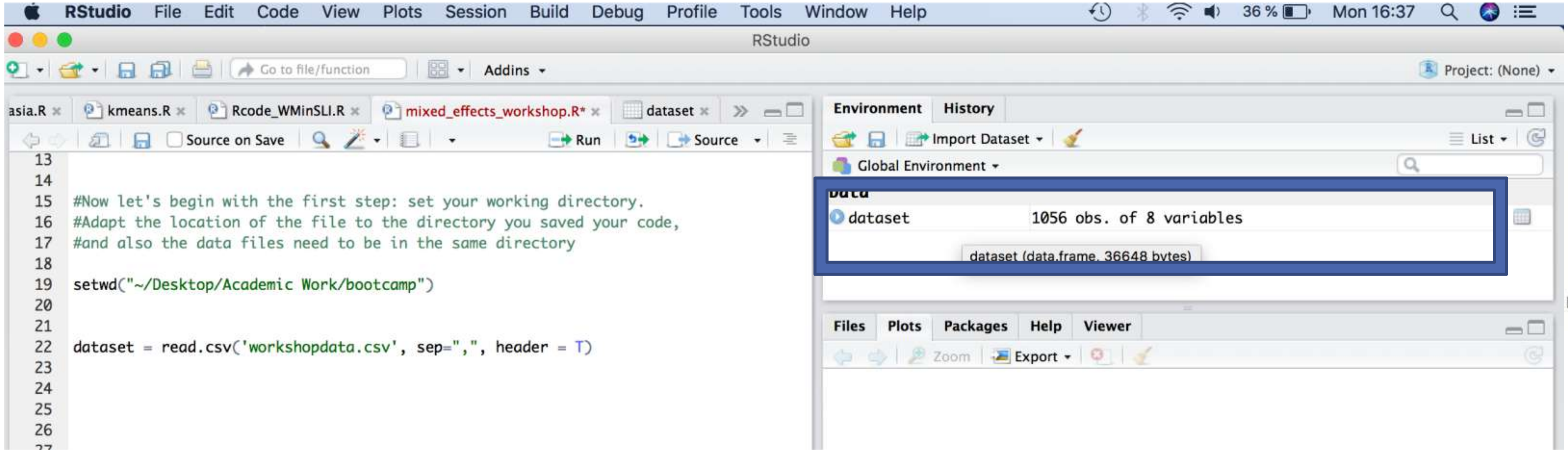
- Keep calm
- Check if the data file
- Check for missing commas or parentheses

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models**

How to import your data into R



International school in psycholinguistics, neurolinguistics and clinical linguistics

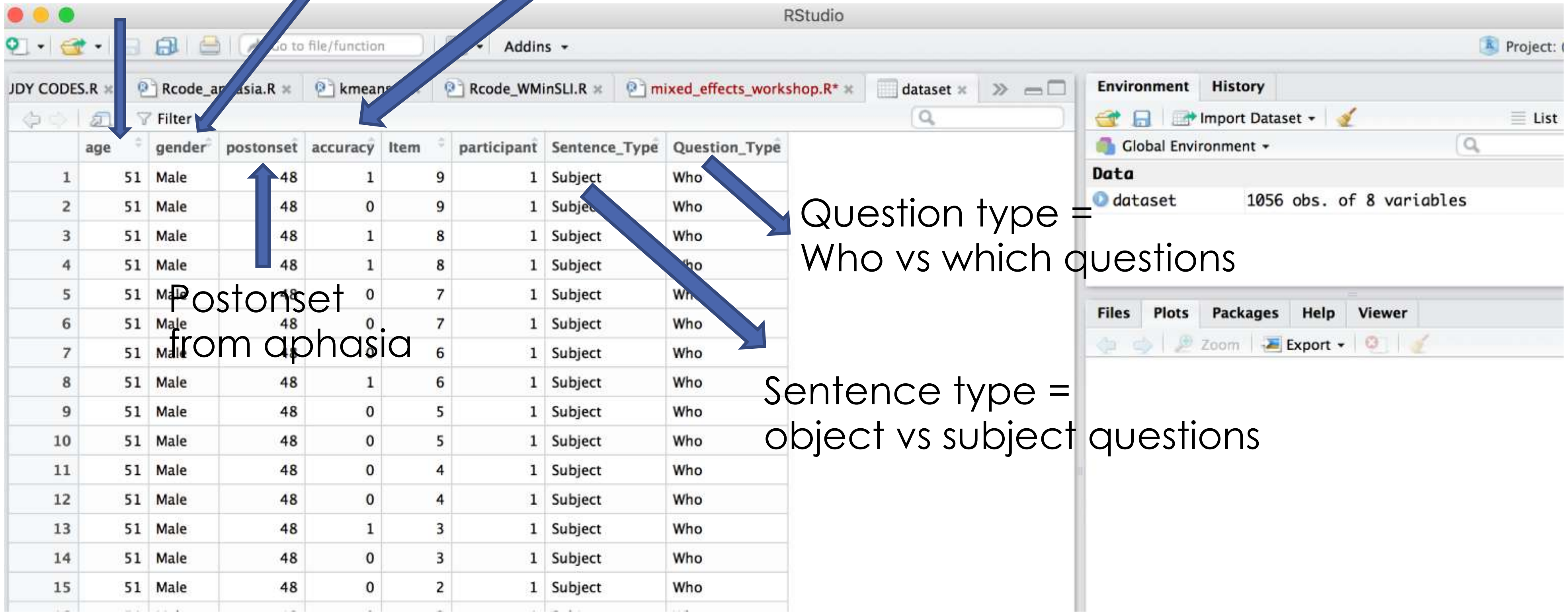
UNIVERSITÉ CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models**

Age of participants

gender of participants

Accuracy (0,1)



	age	gender	postonset	accuracy	Item	participant	Sentence_Type	Question_Type
1	51	Male	48	1	9	1	Subject	Who
2	51	Male	48	0	9	1	Subject	Who
3	51	Male	48	1	8	1	Subject	Who
4	51	Male	48	1	8	1	Subject	Who
5	51	Male	48	0	7	1	Subject	Who
6	51	Male	48	0	7	1	Subject	Who
7	51	Male	[illegible]	[illegible]	6	1	Subject	Who
8	51	Male	48	1	6	1	Subject	Who
9	51	Male	48	0	5	1	Subject	Who
10	51	Male	48	0	5	1	Subject	Who
11	51	Male	48	0	4	1	Subject	Who
12	51	Male	48	0	4	1	Subject	Who
13	51	Male	48	1	3	1	Subject	Who
14	51	Male	48	0	3	1	Subject	Who
15	51	Male	48	0	2	1	Subject	Who

Postonset from aphasia

Question type = Who vs which questions

Sentence type = object vs subject questions

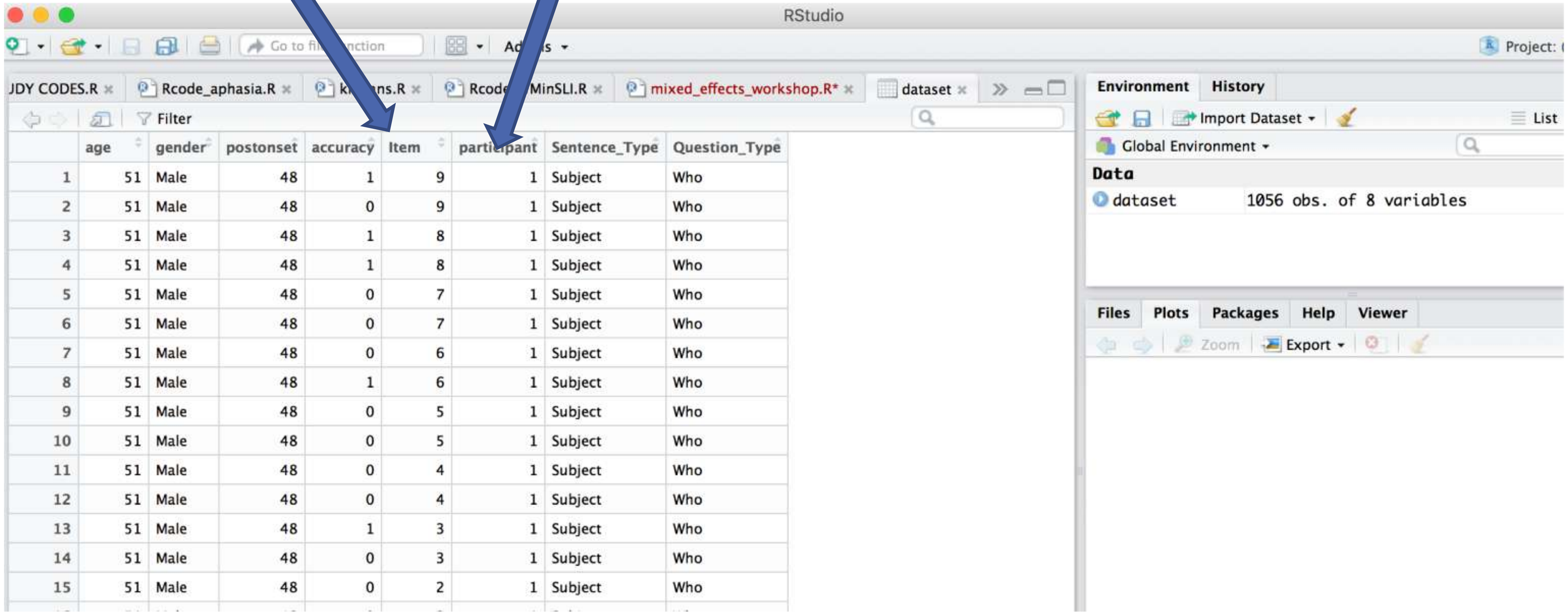
International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models**

Experimental Items

Participants (people with aphasia)



	age	gender	postonset	accuracy	Item	participant	Sentence_Type	Question_Type
1	51	Male	48	1	9	1	Subject	Who
2	51	Male	48	0	9	1	Subject	Who
3	51	Male	48	1	8	1	Subject	Who
4	51	Male	48	1	8	1	Subject	Who
5	51	Male	48	0	7	1	Subject	Who
6	51	Male	48	0	7	1	Subject	Who
7	51	Male	48	0	6	1	Subject	Who
8	51	Male	48	1	6	1	Subject	Who
9	51	Male	48	0	5	1	Subject	Who
10	51	Male	48	0	5	1	Subject	Who
11	51	Male	48	0	4	1	Subject	Who
12	51	Male	48	0	4	1	Subject	Who
13	51	Male	48	1	3	1	Subject	Who
14	51	Male	48	0	3	1	Subject	Who
15	51	Male	48	0	2	1	Subject	Who

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models**

To get an idea about the data,

Run str() function

```
> str(dataset)
'data.frame':	1056 obs. of  8 variables:
 $ age          : int  51 51 51 51 51 51 51 51 51 51 ...
 $ gender       : Factor w/ 2 levels "Female","Male": 2 2 2 2 2 2 2 2 2 2 ...
 $ postonset    : int  48 48 48 48 48 48 48 48 48 48 ...
 $ accuracy     : int  1 0 1 1 0 0 0 1 0 0 ...
 $ Item         : int  9 9 8 8 7 7 6 6 5 5 ...
 $ participant  : int  1 1 1 1 1 1 1 1 1 1 ...
 $ Sentence_Type: Factor w/ 2 levels "Object","Subject": 2 2 2 2 2 2 2 2 2 2 ...
 $ Question_Type: Factor w/ 2 levels "Which","Who": 2 2 2 2 2 2 2 2 2 2 ...
```

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models**

To get an idea about the data,

Run summary() function

```
> summary(dataset)
      age             gender       postonset          accuracy            Item         participant
 Min.   :51.00   Female:192   Min.   :  1.00   Min.   :0.0000   Min.   : 1.00   Min.   : 1
 1st Qu.:56.00   Male  :864   1st Qu.:  1.00   1st Qu.:0.0000   1st Qu.: 3.75   1st Qu.: 3
 Median :63.00                Median : 11.00   Median :1.0000   Median : 6.50   Median : 6
 Mean   :62.82                Mean   : 27.73   Mean   :0.5881   Mean   : 6.50   Mean   : 6
 3rd Qu.:69.00                3rd Qu.: 22.00   3rd Qu.:1.0000   3rd Qu.: 9.25   3rd Qu.: 9
 Max.   :74.00                Max.   :180.00   Max.   :1.0000   Max.   :12.00   Max.   :11
 Sentence_Type Question_Type
 Object :528   Which:528
 Subject:528   Who  :528
```

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models**

To get an idea about the data,

Run head() function

```
> head(dataset)
  age gender postonset accuracy Item participant Sentence_Type Question_Type
1  51   Male        48        1    9           1       Subject           Who
2  51   Male        48        0    9           1       Subject           Who
3  51   Male        48        1    8           1       Subject           Who
4  51   Male        48        1    8           1       Subject           Who
5  51   Male        48        0    7           1       Subject           Who
6  51   Male        48        0    7           1       Subject           Who
> 
```

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Visualisation

boxplot()

hist()

qqnorm() & qqline()

plot()

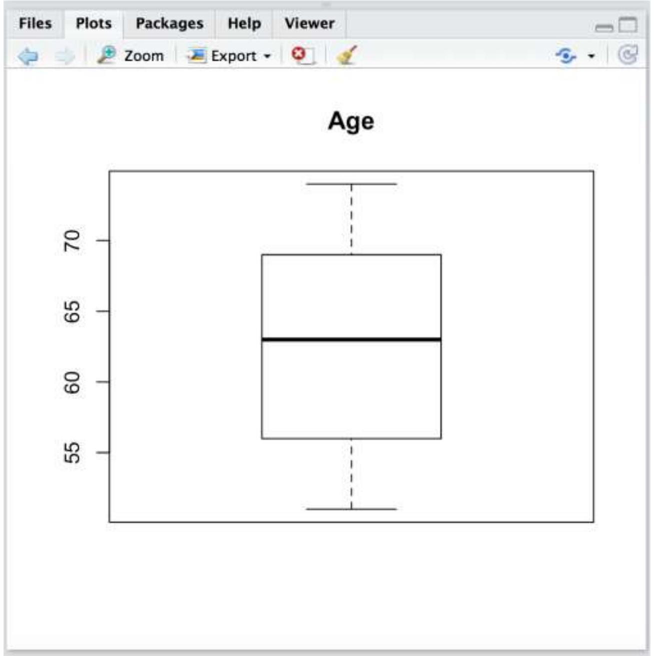
barplot()

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Visualisation

```
boxplot(dataset$age, main = "Age")
```

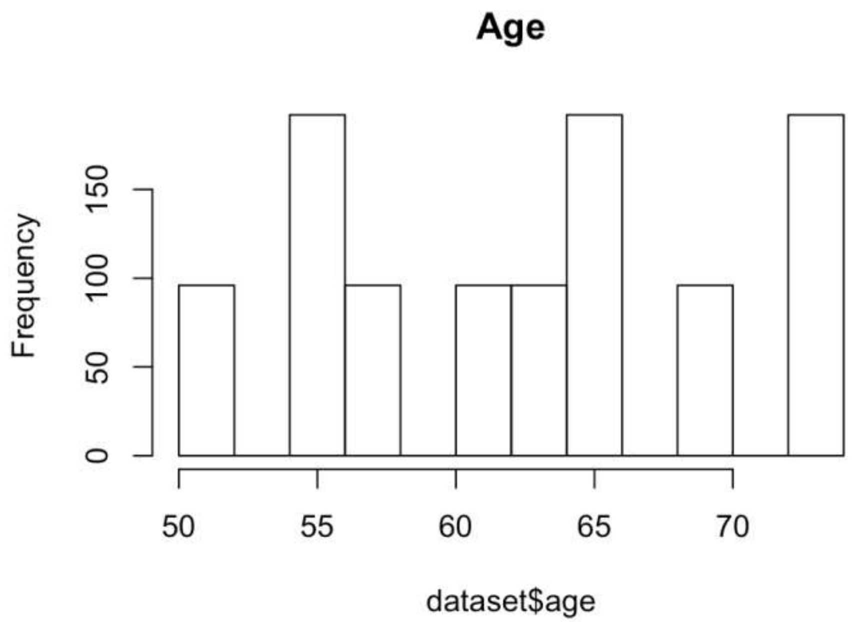


International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Visualisation

hist(dataset$age, main = "Age")



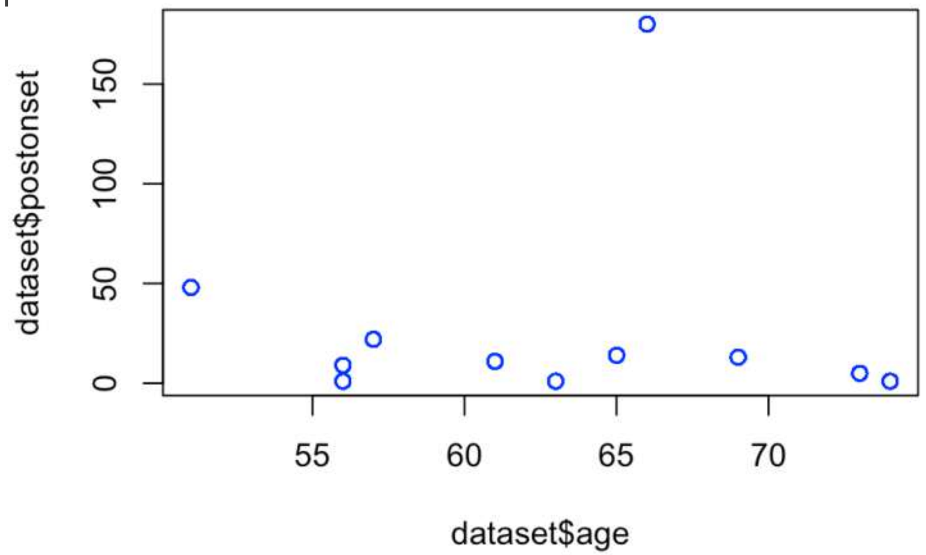
International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Visualisation

plot(dataset$age, dataset$postonset, col="blue")

This will return a simple scatter plot with Postonset and age



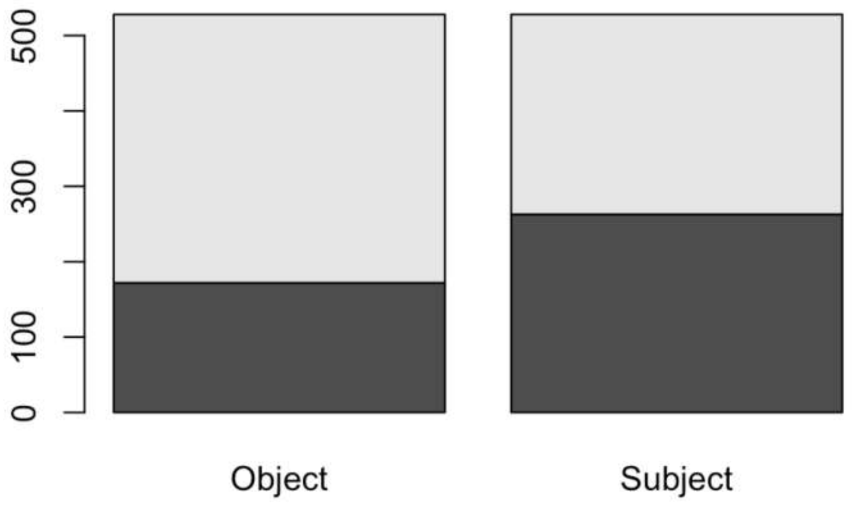
International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Visualisation

plot.table = table (dataset$accuracy, dataset$Sentence_Type)

barplot (plot.table)



International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Visualisation: ggplot

- **A quick peek into the data: visualize the data**

```
install.packages("plyr")
install.packages("ggplot2")

library(plyr)
library(ggplot2)
```

```
aphasia.mean <- ddply(dataset, c( "Sentence_Type", "Question_Type"), summarise,
                    N    = length(accuracy),
                    mean = mean(accuracy,na.rm=TRUE),
                    sd   = sd(accuracy,na.rm=TRUE),
                    se   = sd / sqrt(N))
```

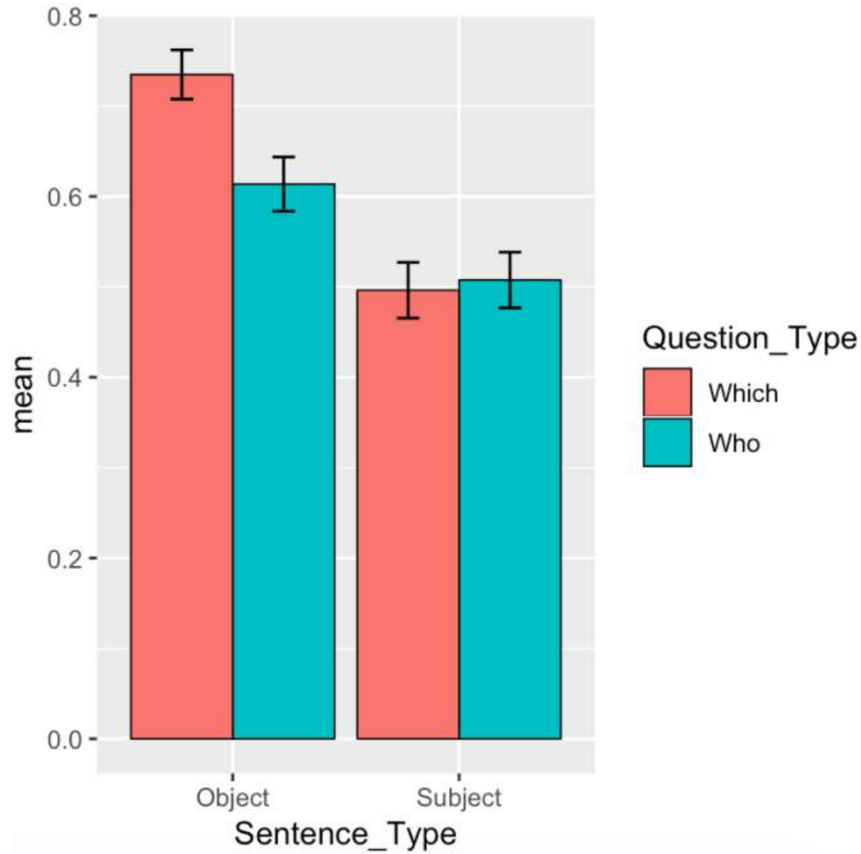
International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Visualisation: ggplot

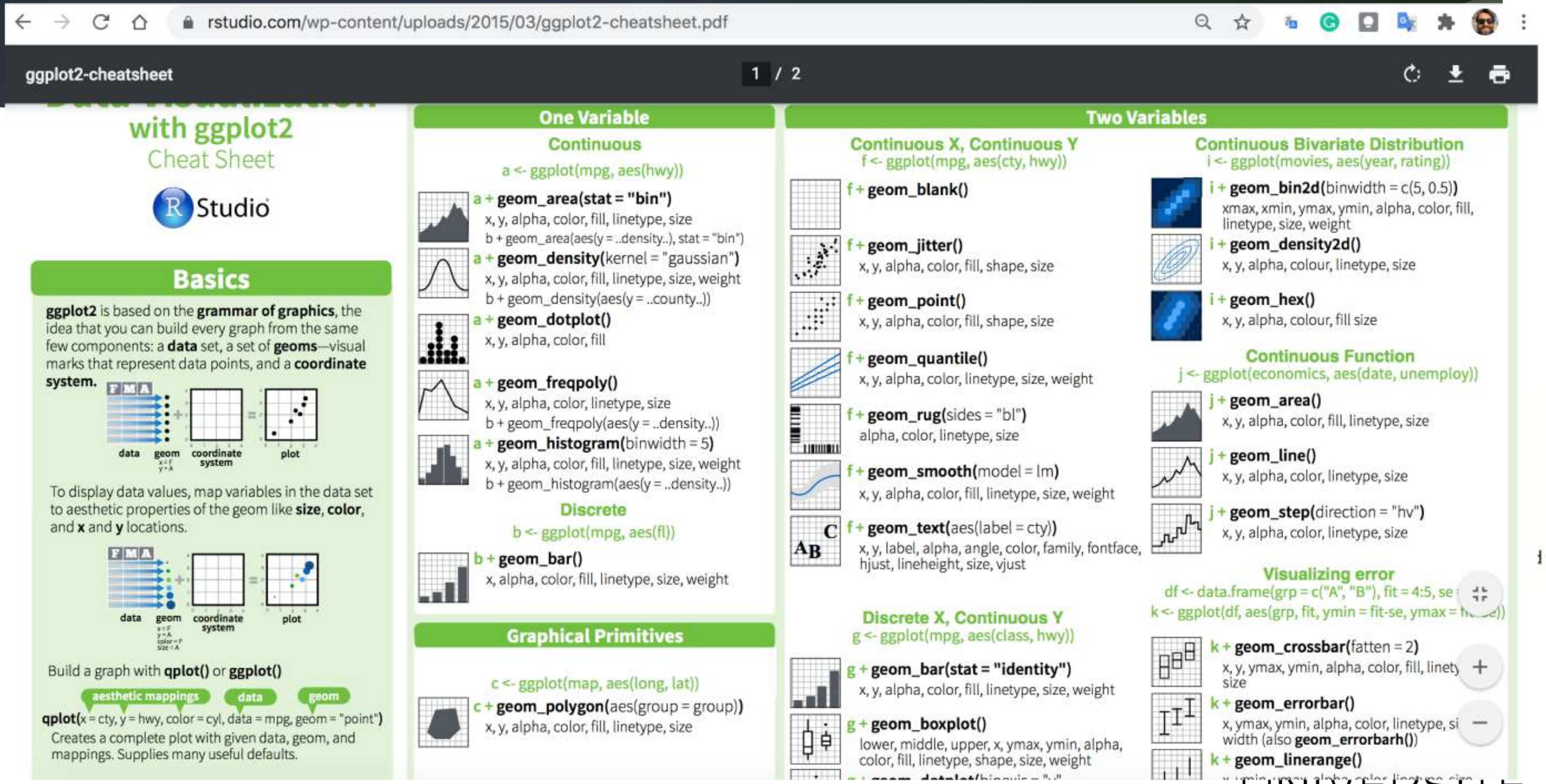
- **A quick peek into the data: visualize the data**

```
ggplot(aphasia.mean, aes(x=Sentence_Type, fill =Question_Type, y=mean, label=mean)) +
  geom_bar(position=position_dodge(), stat="identity", colour="black", size=.3) +
  geom_errorbar(aes(ymin=mean-se, ymax=mean+se),
                width=.2,                    # Width of the error bars
                position=position_dodge(.9))
```



Visualisation

Ggplot cheatsheet



Hitchhikers' guide to **mixed effects regression models** Logistic ME regression

4.1 Understanding mixed and fixed effects

UNIVERSITÉ CÔTE D'AZUR

Dr. Seçkin Arslan

seckin1984@gmail.com

@seckin1984

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

26.10.21

Hitchhikers' guide to **mixed effects regression models**

- **Running a generalized linear mixed-effect regression (glmer) model for accuracy data.**

glmer (dependent var ~ independent var1 * independent

var2…. + (1 | random intercept) + (1 | random intercept),

family = binomial, data = dataset))

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models**

- **Running a generalized linear mixed-effect regression (glmer) model for accuracy data.**

```
install.packages("lme4")
library(lme4)

model1 = glmer(accuracy ~ Sentence_Type * Question_Type  +  (1|participant) + (1|Item), family=binomial, data=dataset)
```

Independent variables

Notice this is glmer

Dependent variable
What we measured

Random intercepts

UNIVERSITÉ CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models**

- **Simple random effect structure**

```
model1 = glmer(accuracy ~ Sentence_Type * Question_Type  +
(1|Item),
          family=binomial, data=dataset)

summary(model1)
```

Fixed-effects

Random effect

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitchh
mixed effec

- **Simple random effect structu**

  # summary (model1

Output ---->

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [
glmerMod]
 Family: binomial  ( logit )
Formula: accuracy ~ Sentence_Type * Question_Type + (1 | Item)
   Data: dataset

     AIC      BIC   logLik deviance df.resid
  1396.0   1420.8   -693.0   1386.0     1051

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.9957 -1.0301  0.5888  0.8454  1.1906

Random effects:
 Groups Name        Variance Std.Dev.
 Item   (Intercept) 0.04853  0.2203
Number of obs: 1056, groups:  Item, 12

Fixed effects:
                                    Estimate Std. Error z value Pr(>|z|)
(Intercept)                           1.0307     0.1540   6.692 2.21e-11 ***
Sentence_TypeSubject                 -1.0460     0.1871  -5.590 2.27e-08 ***
Question_TypeWho                     -0.5626     0.1892  -2.974  0.00294 **
Sentence_TypeSubject:Question_TypeWho  0.6086     0.2578   2.361  0.01825 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) Snt_TS Qst_TW
Sntnc_TypSb -0.683
Qustn_TypWh -0.674  0.555
Snt_TS:Q_TW  0.495 -0.725 -0.734
```

Hitchhikers' guide to **mixed effects regression models**

- **Simple random intercept with nested factors**

model1 = glmer(accuracy ~ Sentence_Type * Question_Type + (1 | participant / gender), family=binomial, data=dataset)

summary(model1)

Fixed-effects

Gender as nested factor within subjects

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitchhi
mixed effec

- **Simple random interce**

model1 = glmer(accura
(1 | participant/gender),

summary(model1)

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [
glmerMod]
 Family: binomial  ( logit )
Formula: accuracy ~ Sentence_Type * Question_Type + (1 | participant/gender)
   Data: dataset

     AIC      BIC   logLik deviance df.resid
  1292.8   1322.6   -640.4   1280.8     1050

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.6140 -0.8180  0.4352  0.7570  1.7272

Random effects:
 Groups             Name        Variance Std.Dev.
 gender:participant (Intercept) 0.08661  0.2943
 participant        (Intercept) 0.58828  0.7670
Number of obs: 1056, groups:  gender:participant, 11; participant, 11

Fixed effects:
                                    Estimate Std. Error z value Pr(>|z|)
(Intercept)                           1.1734     0.2893   4.056 4.98e-05 ***
Sentence_TypeSubject                 -1.1770     0.1991  -5.911 3.40e-09 ***
Question_TypeWho                     -0.6269     0.1996  -3.140  0.00169 **
Sentence_TypeSubject:Question_TypeWho  0.6796     0.2737   2.483  0.01302 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) Snt_TS Qst_TW
Sntnc_TypSb -0.385
Qustn_TypWh -0.378  0.549
Snt_TS:Q_TW  0.276 -0.722 -0.730
```

Hitchhikers' guide to **mixed effects regression models**

- **Two independent random intercepts**

model3 = glmer(accuracy ~ Sentence_Type * Question_Type +
        (1 | Item) + (1 | participant), family=binomial, data=dataset)

summary(model3)

Fixed-effects

Two independent random effects

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitch
mixed effe

- **Two independent ro**

model3 = glmer(acc
 (1 | Item) + (1 |

summary(model3)

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [
glmerMod]
 Family: binomial  ( logit )
Formula: accuracy ~ Sentence_Type * Question_Type + (1 | Item) + (1 |
    participant)
   Data: dataset

     AIC      BIC   logLik deviance df.resid
  1286.8   1316.6   -637.4   1274.8     1050

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.7514 -0.8095  0.4149  0.7509  1.9445

Random effects:
 Groups      Name        Variance Std.Dev.
 Item        (Intercept) 0.08219  0.2867
 participant (Intercept) 0.70366  0.8388
Number of obs: 1056, groups:  Item, 12; participant, 11

Fixed effects:
                                     Estimate Std. Error z value Pr(>|z|)
(Intercept)                            1.1927     0.3059   3.899 9.65e-05 ***
Sentence_TypeSubject                  -1.1965     0.2009  -5.957 2.57e-09 ***
Question_TypeWho                      -0.6372     0.2012  -3.167  0.00154 **
Sentence_TypeSubject:Question_TypeWho  0.6907     0.2758   2.504  0.01227 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) Snt_TS Qst_TW
Sntnc_TypSb -0.368
Qustn_TypWh -0.361  0.549
```

Hitchhikers' guide to **mixed effects regression models**

- **Random slope**

```
model4 = glmer(accuracy ~ Sentence_Type * Question_Type  +
            (age | Item) + (1 | participant), family=binomial,
data=dataset)

summary(model4)
```

Fixed-effects

Random slopes for age per item

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitch

mixed effe

- **Random slope**

```
model4 = glmer(acc
           (age | Item
data=dataset)

summary(model4)
```

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [
glmerMod]
 Family: binomial  ( logit )
Formula: accuracy ~ Sentence_Type * Question_Type + (age | Item) + (1 |
    participant)
   Data: dataset

     AIC      BIC   logLik deviance df.resid
  1289.8   1329.5   -636.9   1273.8     1048

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.7401 -0.8085  0.4130  0.7509  1.8623

Random effects:
 Groups      Name        Variance  Std.Dev. Corr
 Item        (Intercept) 1.0139156 1.00693
             age         0.0001293 0.01137  -1.00
 participant (Intercept) 0.7084785 0.84171
Number of obs: 1056, groups:  Item, 12; participant, 11

Fixed effects:
                                      Estimate Std. Error z value Pr(>|z|)
(Intercept)                             1.1974     0.3071   3.899 9.67e-05 ***
Sentence_TypeSubject                   -1.1990     0.2011  -5.962 2.49e-09 ***
Question_TypeWho                       -0.6389     0.2015  -3.171  0.00152 **
Sentence_TypeSubject:Question_TypeWho   0.6924     0.2761   2.508  0.01213 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) Snt_TS Qst_TW
Sntnc_TypSb -0.367
```

Hitchhikers' guide to **mixed effects regression models**

- **Adding a fixed-effect**

```
model5 = glmer(accuracy ~ Sentence_Type * Question_Type  +
postonset +
            (1 | Item) + (1 | participant), family=binomial,
data=dataset)


summary(model5)
```

Fixed-effects

Random slopes for age per item

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitch
mixed effe

- **Adding a fixed-effe**

model5 = glmer(acc
postonset +
(1 | Item) +
data=dataset)

summary(model5)

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [
glmerMod]
 Family: binomial  ( logit )
Formula: accuracy ~ Sentence_Type * Question_Type + postonset + (1 | Item) +
    (1 | participant)
   Data: dataset

     AIC      BIC   logLik deviance df.resid
  1286.9   1321.6   -636.5   1272.9     1049

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.7377 -0.8135  0.4142  0.7525  1.9214

Random effects:
 Groups      Name        Variance Std.Dev.
 Item        (Intercept) 0.08189  0.2862
 participant (Intercept) 0.58359  0.7639
Number of obs: 1056, groups:  Item, 12; participant, 11

Fixed effects:
                                       Estimate Std. Error z value Pr(>|z|)
(Intercept)                            1.383913   0.318443   4.346 1.39e-05 ***
Sentence_TypeSubject                  -1.196587   0.200882  -5.957 2.57e-09 ***
Question_TypeWho                      -0.637187   0.201207  -3.167  0.00154 **
postonset                             -0.006901   0.004809  -1.435  0.15123
Sentence_TypeSubject:Question_TypeWho  0.690730   0.275826   2.504  0.01227 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) Snt_TS Qst_TW pstnst
Sntnc_TypSb -0.360
```

Hitchhikers' guide to **mixed effects regression models**

- **Adding a fixed-effect**

```
model6 = glmer(accuracy ~ Sentence_Type * Question_Type  + postonset +
          (postonset | Item) , family=binomial, data=dataset)

summary(model6)
```

Fixed-effects

Random slopes for age per item

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitc
mixed eff

- **Adding a fixed-ef**

model6 = glmer(a
postonset +

(postons

summary(model6)

```
> summary(model6)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [
glmerMod]
 Family: binomial  ( logit )
Formula: accuracy ~ Sentence_Type * Question_Type + postonset + (postonset |
    Item)
   Data: dataset

     AIC      BIC   logLik deviance df.resid
  1378.8   1418.5   -681.4   1362.8     1048

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.1710 -1.0449  0.5558  0.8758  1.7656

Random effects:
 Groups Name        Variance  Std.Dev. Corr
 Item   (Intercept) 4.882e-02 0.220949
        postonset   6.867e-08 0.000262 1.00
Number of obs: 1056, groups:  Item, 12

Fixed effects:
                                       Estimate Std. Error z value Pr(>|z|)
(Intercept)                            1.221902   0.161690   7.557 4.12e-14 ***
Sentence_TypeSubject                  -1.071318   0.189616  -5.650 1.61e-08 ***
Question_TypeWho                      -0.576685   0.191617  -3.010  0.00262 **
postonset                             -0.006213   0.001323  -4.694 2.67e-06 ***
Sentence_TypeSubject:Question_TypeWho  0.623672   0.260887   2.391  0.01682 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) Snt_TS Qst_TW pstnst
```

Hitchhikers' guide to **mixed effects regression models**

Additional consideration

When specifying the random effect structure we use (1|Item)

Here 1 refers to the "intercept" - (1|Item) is an intercept only model (i.e. slopes per item do not vary)

Alternatives:

(Age|Item) – slopes per item vary by age

(1 +Age |Item) – slopes per item vary by age plus the intercept

(0+Age|Item) – slopes per item vary by age we force to exclude intercept

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models**

- **Which model do we choose?**

- Prefer more parsimonious models
- Rule of thumb: "**smaller the better**" - chose models with smaller AIC/BIC values
- Compare models:

anova(model1, model2)

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models** **Logistic ME regression**

4.3. Interpreting logistic regression

UNIVERSITÉ CÔTE D'AZUR

Dr. Seçkin Arslan

seckin1984@gmail.com

@seckin1984

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

26.10.21

Hitchhikers' guide to **mixed effects regression models**

- **Interpreting logistic coefficients**

#fixef(model1)

#plogis(fixef(model1) ["(Intercept)"])

= 0.7672304

On average 76% chance of being correct

International school in psycholinguistics, neurolinguistics and clinical linguistics

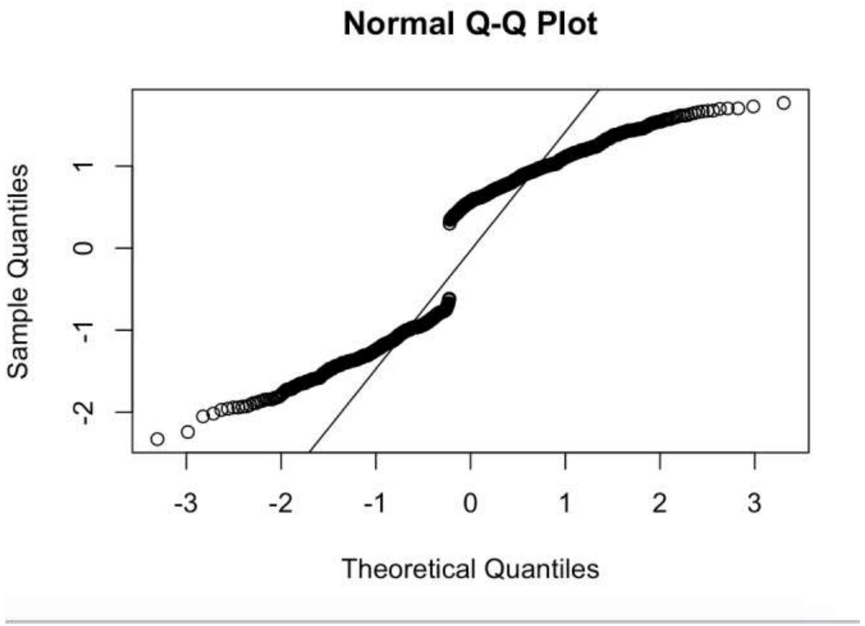
UNIVERSITÉ CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models**

- **Inspecting distribution of your residuals**

qqnorm (resid(model1))

qqline (resid(model1))



International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models**

Post-hoc tests with emmeans

emmeans(model4, "Question_Type", by = "Sentence_Type")

pairs(emmeans(model4, "Question_Type", by = "Sentence_Type"))

```
> pairs(emmeans(model4, "Question_Type", by = "Sentence_Type"))
Sentence_Type = Object:
 contrast    estimate    SE  df z.ratio p.value
 Which - Who   0.6389 0.201 Inf   3.171  0.0015

Sentence_Type = Subject:
 contrast    estimate    SE  df z.ratio p.value
 Which - Who  -0.0536 0.189 Inf  -0.284  0.7763

Results are given on the log odds ratio (not the response) scale.
```

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models**

Alternative post-hocs

#first subset your data into different question types

data.who <- subset(dataset, Question_Type == 'Who')

data.which <- subset(dataset, Question_Type == 'Which')

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models**

library(multcomp) #Install.package("multcomp") first if you don't have this package

data.who$Factor = interaction(data.who$Sentence_Type)

model = glmer(accuracy ~ Factor + (1|Item) + (1|participant), data=data.who, family = binomial)

summary(model)

summary(glht(model,linfct=mcp(Factor = "Tukey")))

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models**

```
Linear Hypotheses:
                        Estimate Std. Error z value Pr(>|z|)
Subject - Object == 0    -0.5199     0.1936  -2.686  0.00724 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

Great subject – object sentences in who questions are different…

International school in
cholinguistics, neurolinguistics and
clinical linguistics

UNIVERSITÉ
CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models**

How should you report a logistic mixed-effects regression model?

A best way is to provide a table

Fixed effects:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	1.383913	0.318443	4.346	1.39e-05	***
Sentence_TypeSubject	-1.196587	0.200883	-5.957	2.57e-09	***
Question_TypeWho	-0.637187	0.201208	-3.167	0.00154	**
postonset	-0.006901	0.004809	-1.435	0.15123	
Sentence_TypeSubject:Question_TypeWho	0.690730	0.275828	2.504	0.01227	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitchhikers' guide to **mixed effects regression models**

How should you report a logistic mixed-effects regression model?

"outputs from a generalized logistic mixed-effects regression model have shown significant effects of Question Type and Sentence Type and of interactions between the two." (see table X). s

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR

Hitchhikers'
mixed effects reg

```
> pairs(emmeans(model4, "Question_Type", by = "Sentence_Type"))
Sentence_Type = Object:
 contrast     estimate    SE  df z.ratio p.value
 Which - Who    0.6389 0.201 Inf   3.171  0.0015

Sentence_Type = Subject:
 contrast     estimate    SE  df z.ratio p.value
 Which - Who   -0.0536 0.189 Inf  -0.284  0.7763

Results are given on the log odds ratio (not the response) scale.
```

How should you report a post-hoc test results

"A series of post-hoc tests confirmed that the subjects were more accurate in responding to "which questions" than "who questions" (ß = 0.63, SE = 0.20, z = 3.17).

International school in psycholinguistics, neurolinguistics and clinical linguistics

UNIVERSITÉ CÔTE D'AZUR